ACADEMIC PRESS

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT

Journal of Multivariate Analysis 88 (2004) 109–117

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

A chi-square test for dimensionality with non-Gaussian data

Z.D. Bai[a,b,*,1] and Xuming He[c,2]

[a] *Department of Mathematics, Northeast Normal University, Changchun 130024, China*
[b] *Department of Statistics and Applied Probability, National University of Singapore, Singapore 119260, Singapore*
[c] *Department of Statistics, University of Illinois at Urbana-Champaign, IL, USA*

Received 2 September 2001

Abstract

The classical theory for testing the null hypothesis that a set of canonical correlation coefficients is zero leads to a chi-square test under the assumption of multi-normality. The test has been used in the context of dimension reduction. In this paper, we study the limiting distribution of the test statistic without the normality assumption, and obtain a necessary and sufficient condition for the chi-square limiting distribution to hold. Implications of the result are also discussed for the problem of dimension reduction.

AMS 2000 subject classifications: primary 62H20; secondary 62J02; 62E20

Keywords: Canonical correlation; Chi-square test; Dimension reduction; Inverse regression; SIR models

1. Introduction

An important class of dimension reduction techniques assumes that the relationship between $\mathbf{x} \in R^p$ and $y \in R$ can be described by

$$y = g(\mathbf{x}'\beta_1, \ldots, \mathbf{x}'\beta_K, \varepsilon), \tag{1}$$

*Corresponding author.
E-mail address: stabaizd@leonis.nus.edu.sg (Z.D. Bai).
[1] Research supported in part by Research Grant R-155-000-014-112 from the National University of Singapore and in part by NSFC 201471000.
[2] Research supported in part by NSF Grant DMS-0102411. The authors wish to thank Professor W.K. Fung for fruitful discussions.

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00056-3

where $\beta_1, \dots, \beta_K$ span a K-dimensional subspace with $K \leqslant p$, ε is independent of $\mathbf{x}$, and g is an unspecified function. When p is relatively large and K is small, the objective of dimension reduction is achieved by finding the K-dimensional subspace of $\mathbf{x}$.

Li [10] and Duan and Li [5] laid the foundation for sliced inverse regression (SIR) as a tool of estimating the subspace spanned by β_j's. Subsequent work by Cook [3] provided a supplementary method using the sliced average variance estimation (SAVE). He and Shen [9] and Fung et al. [8] proposed CANCOR, a variant of SIR that uses canonical correlation between $\mathbf{x}$ and a spline basis for y. We refer to a recent discussion paper by Cook and Yin [4] for a general framework for understanding dimension reduction.

One practical question associated with those methods is to determine the necessary number of dimensions K. Under model (1) and under a linearity condition given in [10], the rank of the matrix $Cov(E(\mathbf{x}|y))$ is at most K. Li [10] proposed a chi-square test based on the eigenvalues of an estimated matrix of $Cov(E(\mathbf{x}|y))$. An equivalent test may be illustrated using CANCOR as follows.

Following Fung et al. [8], we assume without loss of generality that $y \in [0, 1]$ and let $\pi(y) \in R^{q+1}$ be a B-spline basis function on the support of y, where q is the number of linearly independent basis functions in π. Consider the nontrivial case with $q > K$. Let $\hat{\lambda}_1 \geqslant \cdots \geqslant \hat{\lambda}_r$ be the canonical correlation coefficients between $\{\mathbf{x}_i\}$ and $\{\pi(y_i)\}$, where $r = \min(p, q)$. As shown by Fung et al. [8], the rank of $Cov(E(\mathbf{x}|y))$ is equal to the number of nonzero (population) canonical correlation coefficients. To test the null hypothesis that the rank of $Cov(E(\mathbf{x}|y))$ is K so that the last $p - K$ canonical correlation coefficients are zero, we may use the following test statistic:

$$T_n = -(n - (p + q + 3)/2) \sum_{j=K+1}^{r} \log(1 - \hat{\lambda}_j^2) \tag{2}$$

and reject the null hypothesis at level α if $T_n > \chi^2_{(p-K)(q-K),\alpha}$. The basic form of the test statistic and Bartlett's modification used in (2) can be found in [1, p. 498]. If $\mathbf{x}$ is normally distributed, Li [10, Theorem 5.1] showed in the context of SIR that T_n is asymptotically chi-square distributed. Fung et al. [8] used this test in a simulation comparison with some other tests and found that the chi-square test is often reliable under model (1) even when $\mathbf{x}$ is not normal provided that the distribution of $\mathbf{x}$ is not highly skewed or heavy-tailed. The purpose of this note is to provide general assumptions under which the chi-square test is asymptotically valid for model (1). In some cases, the normality of $\mathbf{x}$ is not needed to apply the chi-square test. Our main result is given in Section 2 but some convergence results on singular value decompositions and their proofs are provided in Section 3. The techniques used in Section 3 can be viewed as an extension of those in [6].

2. Main result

For notational simplicity, we shall use $(\mathbf{x}_i, \mathbf{y}_i)$ $i = 1, 2, \ldots, n$ as random vectors in R^{p+q}. We can replace $\mathbf{y}$ by $\pi(y)$ for the purpose of applying our result to the CANCOR method described above.

Thanks to the affine invariance property of canonical correlation coefficients, we can assume without loss of generality that

$$\mathbf{Cov}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \Sigma \\ \Sigma' & \mathbf{I}_q \end{pmatrix},$$

where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix, and Σ is a $p \times q$ diagonal matrix with only K positive canonical correlations on the north-west corner. The affine transformation used on $\mathbf{x}$ enables us to split $\mathbf{x}$ as $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')' \in R^K \times R^{p-K}$ so that $\mathbf{y}$ is uncorrelated with $\mathbf{x}_2$ under the null hypothesis. This corresponds to the case $(\beta_1, \ldots, \beta_K) = (\mathbf{I}_K, 0)'$ in model (1).

First, we consider the matrices $\mathbf{S}_x = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'$, $\mathbf{S}_y = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i\mathbf{y}_i'$, $\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{y}_i'$. By the central limit theorem, we have

$$\begin{pmatrix} \mathbf{S}_x & \mathbf{S}_n \\ \mathbf{S}_n' & \mathbf{S}_y \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \Sigma \\ \Sigma' & \mathbf{I}_q \end{pmatrix} + \frac{1}{\sqrt{n}}\begin{pmatrix} \mathbf{R}_{nx} & \mathbf{R}_n \\ \mathbf{R}_n' & \mathbf{R}_{ny} \end{pmatrix}, \tag{3}$$

with $(\mathbf{R}_{nx}, \mathbf{R}_{ny}, \mathbf{R}_n) \xrightarrow{d.} (\mathbf{R}_x, \mathbf{R}_y, \mathbf{R})$ in distribution, where $(\mathbf{R}_x, \mathbf{R}_y, \mathbf{R})$ is jointly distributed as three matrices of normal entries with mean 0 and variance–covariance structure determined by the variance–covariances of $\mathbf{x}\otimes\mathbf{x}$, $\mathbf{y}\otimes\mathbf{y}$ and $\mathbf{x}\otimes\mathbf{y}$, where $\otimes$ denotes the Kronecker product. By Skorohod theorem, we may assume $(\mathbf{R}_{nx}, \mathbf{R}_{ny}, \mathbf{R}_n) \to (\mathbf{R}_x, \mathbf{R}_y, \mathbf{R})$ holds almost surely. That is

$$\begin{pmatrix} \mathbf{S}_x & \mathbf{S}_n \\ \mathbf{S}_n' & \mathbf{S}_y \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \Sigma \\ \Sigma' & \mathbf{I}_q \end{pmatrix} + \frac{1}{\sqrt{n}}\begin{pmatrix} \mathbf{R}_x & \mathbf{R} \\ \mathbf{R}' & \mathbf{R}_y \end{pmatrix} + o(1/\sqrt{n}). \tag{4}$$

The sample canonical correlations between $\mathbf{x}$ and $\mathbf{y}$ are the singular values of $\mathbf{S}_x^{-1/2}\mathbf{S}_n\mathbf{S}_y^{-1/2}$. By Taylor expansion, it follows from (4) that

$$\mathbf{S}_x^{-1/2}\mathbf{S}_n\mathbf{S}_y^{-1/2} = \Sigma + \frac{1}{\sqrt{n}}\left[\mathbf{R} - \frac{1}{2}(\mathbf{R}_x\Sigma + \Sigma\mathbf{R}_y)\right] + o(1/\sqrt{n}).$$

Now, suppose that the multiplicities of the population canonical correlations are $p_1, \ldots, p_s$, i.e., the matrix Σ has the form $\Sigma = (diag\{\lambda_1\mathbf{I}_{p_1}, \ldots, \lambda_s\mathbf{I}_{p_s}, 0_{p-K,q-K}\})$ with $\lambda_1 > \cdots > \lambda_s > 0$, $p_1 + \cdots + p_s = K$ and $0_{p-K,q-K}$ being a zero matrix of order $(p-K)\times(q-K)$.

Consider the singular value decomposition of $\mathbf{S}_x^{-1/2}\mathbf{S}_n\mathbf{S}_y^{-1/2}$ as

$$\mathbf{S}_x^{-1/2}\mathbf{S}_n\mathbf{S}_y^{-1/2} = \mathbf{U}_n\mathbf{L}_n\mathbf{V}_n',$$

where $r = \min(p, q)$, $\mathbf{L}_n$ is a $r \times r$ diagonal matrix with nonnegative and decreasing diagonal elements, $\mathbf{U}_n$ is a $p \times r$ orthonormal matrix and $\mathbf{V}_n$ is a $q \times r$ orthonormal matrix, that is, $\mathbf{U}_n'\mathbf{U}_n = \mathbf{V}_n'\mathbf{V}_n = \mathbf{I}_r$.

Based on the block diagonals of Σ, we split the matrix

$$\mathbf{L}_n = diag[\mathbf{L}_{n,1}, \dots, \mathbf{L}_{n,s}, \mathbf{L}_{n,s+1}],$$

where $\mathbf{L}_{n,j}$ is a $p_j \times p_j$ matrix for $j = 1, \dots, s+1$ with $p_{s+1} = r - K$.

By Lemma 3 given in Section 3, we have, under the null hypothesis, that the diagonal elements of $\sqrt{n}\mathbf{L}_{s+1}$ tend to the set of singular values of the $(p-K) \times (q-K)$ lower-right submatrix of $\mathbf{R} - \frac{1}{2}(\mathbf{R}_x\Sigma + \Sigma\mathbf{R}_y)$, which is the same as the $(p-K) \times (q-K)$ lower-right submatrix of $\mathbf{R}$ since the corresponding submatrices of $\mathbf{R}_x\Sigma$ and $\Sigma\mathbf{R}_y$ are both 0.

To get the distribution of the $(p-K) \times (q-K)$ lower-right submatrix of $\mathbf{R}$, note that

$$\mathbf{Cov}(vec(\mathbf{R}')) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

where $\Sigma_1 = \mathbf{Cov}(\mathbf{x}_1 \otimes \mathbf{y})$, $\Sigma_2 = \mathbf{Cov}(\mathbf{x}_2 \otimes \mathbf{y})$, and $vec(\mathbf{R})$ denotes the vectorization of the matrix R (by stacking up the columns of $\mathbf{R}$). This shows that the entries of the $(p-K) \times (q-K)$ lower-right submatrix of $\mathbf{R}$ are jointly normally distributed with mean 0. Consider the trace of $\mathbf{R}'\mathbf{R}$ as a quadratic function of the $(p-K)(q-K)$ normal variables. By Theorem 2 on page 57 of [12], the trace of $\mathbf{R}'\mathbf{R}$ has a chi-square distribution with $(p-K)(q-K)$ degrees of freedom if and only if $\Sigma_2 = \mathbf{I}_{(p-K)(q-K)}$. In our setting, this condition is equivalent to $\mathbf{x}_2$ and $\mathbf{y}$ being *second-order uncorrelated*, that is, for any components a_1, a_2 of $\mathbf{x}_2$ and any components b_1, b_2 of $\mathbf{y}$, the equality $E(a_1b_1a_2b_2) = E(a_1a_2)E(b_1b_2)$ holds.

Let $\hat{\lambda}_j$ $(K+1 \leqslant j \leqslant p)$ be the diagonal elements of $\mathbf{L}_{n,s+1}$. We have obtained our main result as follows.

Theorem 1. *Suppose that* $\mathbf{Cov}(\mathbf{x} \otimes \mathbf{y})$ *exists. Then under model* (1) *the necessary and sufficient condition for* $T_n = n\sum_{j=K+1}^{p} \hat{\lambda}_j^2 \to \chi^2_{(p-K)(q-K)}$ *as* $n \to \infty$ *is that* $\mathbf{x}_2$ *and* $\mathbf{y}$ *are second-order uncorrelated in the sense described above.*

Remark 1. In Section 3, we prove Lemma 3 which also implies that the diagonal elements of $\sqrt{n}(\mathbf{L}_{nj} - \lambda_j\mathbf{I}_{p_j})$ converge to the set of eigenvalues of the $p_j \times p_j$ symmetric matrix $\frac{1}{2}(\mathbf{R}_{jj} + \mathbf{R}_{jj}^* - \lambda_j(\mathbf{R}_{x,jj} + \mathbf{R}_{y,jj}))$, where $\mathbf{R}_{jj}$ denotes the $p_j \times p_j$ diagonal block of $\mathbf{R}$. It also implies the convergence of the eigenmatrices $\mathbf{U}_n$ and $\mathbf{V}_n$. We omit the details here.

Note that $\mathbf{x}_2$ is always orthogonal to $\mathbf{x}_1$ by construction, the conditions of Theorem 1 holds automatically if $\mathbf{x}$ is normally distributed. Also note that the statistic T_n in Theorem 1 is asymptotically equivalent to (2) of Section 1. It may appear that Theorem 1 applies to CANCOR of [8] but not directly to SIR of [10]. However, SIR is equivalent to the canonical correlation approach when the spline basis $\pi(y)$ is replaced by a set of indicator functions. Therefore, Theorem 5.1 of [10] can be viewed as a special case of Theorem 1 with normally distributed $\mathbf{x}$.

The condition of second-order uncorrelatedness holds as long as y is independent of $\mathbf{x}_2$ even if $\mathbf{x}$ is not normal. On the other hand, if we apply Theorem 1 to the test of dimension reduction in CANCOR, $\mathbf{y}$ is formed by spline basis functions for y. For the second-order uncorrelatedness to hold for a sequence of such basis functions, we essentially require y to be independent of $\mathbf{x}_2$. This implies that the strict validity of the chi-square test for dimension reduction is quite limited.

In general, the limiting distribution of T_n is a mixture of chi-squares. The mixture coefficients depend on the eigenvalues of Σ_2, which is not directly estimable. Therefore a transformation of the variables $\mathbf{x}$ towards normality is desirable before using the chi-square test.

There are several other issues that are worth further discussions.

First, the chi-square test is valid only when the first K singular values of Σ are nonzero but the rest are zero. If in the null hypothesis the value K is chosen to be larger than the rank of Σ, the chi-square limiting distribution no longer holds and the test becomes conservative. Therefore, it is recommended that one conducts sequential tests starting from the smallest possible K.

Second, we note that our study assumes that the dimension q is fixed. In theory, this could tend to infinity with n. In the latter case, all we need is to check (3) to make sure that the limiting matrix $\mathbf{R}$ has the same property as in the case of finite q. The limiting distribution of the test statistic would be normal upon a proper standardization. The work by Portnoy [11] provided a means to handle such problems. We also refer to [10, Remark 5.4] for a similar discussion.

Third, the chi-square approximation given by Theorem 1 may not be accurate in some small-sample problems. The simulation study of Fung et al. [8] indicates that in finite-sample problems the chi-square approximation is less accurate when $\mathbf{x}$ is more skewed or heavier-tailed. Cook and Yin [4] proposed a permutation test for the same null hypothesis without explicit moment assumptions, but it has to use the estimated dimension reduction space so its asymptotic distributional theory is not yet available. On the other hand, the chi-square test discussed here does not involve estimating β_j.

Finally, note that the chi-square test is not consistent for all alternatives to model (1). The number of nonzero canonical correlations or the rank of Σ is in general less than or equal to K of model (1) under a linearity condition on the conditional mean of $\mathbf{x}$ given $(\mathbf{x}'\beta_1, \ldots, \mathbf{x}'\beta_K)$. The test T_n concerns the rank of Σ, so it could fail to recover all directions needed in model (1).

3. Lemmas

The proof of Theorem 1 is based on Lemma 3 below. Lemmas 1 and 2 are used as prelude to Lemma 3 in this section.

Lemma 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times q$ $(p \leqslant q)$ *matrices with singular values* a_i *and* b_i $(i = 1, \ldots, p)$, *arranged in descending order, respectively. Then*

$$\sum_{i=1}^{p} (a_i - b_i)^2 \leqslant trace\{(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^*\},$$

where $\mathbf{A}^*$ *denotes the complex conjugate transpose of any matrix* $\mathbf{A}$.

Proof. We have $\sum_i a_i^2 = trace(\mathbf{AA}^*)$ and $\sum_i b_i^2 = trace(\mathbf{BB}^*)$. By Fan [7], we also have

$$\sum_i a_i b_{p-i+1} \leqslant Real\{trace(\mathbf{AB}^*)\} \leqslant \sum_i a_i b_i.$$

Therefore, we get

$$\begin{aligned} \sum_i (a_i - b_i)^2 &= \sum_i a_i^2 + \sum_i b_i^2 - 2\sum_i a_i b_i \\ &\leqslant trace(\mathbf{AA}^*) + trace(\mathbf{BB}^*) - \{trace(\mathbf{AB}^*) + trace(\mathbf{BA}^*)\} \\ &= trace((\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^*). \qquad \square \end{aligned}$$

A similar result can be found in Lemma 2.7 of [2, p. 614] with more details about Fan's inequality.

Lemma 2. *Let* $\mathbf{A}$ *and* $\mathbf{A}_n$ *be* $p \times q$ *complex matrices satisfying*

(1) $\mathbf{A}$ *has rank* r *with positive singular values* $l_1 > \cdots > l_r$.
(2) $\mathbf{A}_n \to \mathbf{A}$ *as* $n \to \infty$;
(3) *Let* $\mathbf{A} = \mathbf{ULV}^*$ *be the singular value decomposition where* $\mathbf{L} = diag\{l_1, \ldots, l_r\}$, $\mathbf{U} \in R^{p \times r}$ *and* $\mathbf{V} \in R^{q \times r}$ *are matrices satisfying* $\mathbf{U}^*\mathbf{U} = \mathbf{V}^*\mathbf{V} = \mathbf{I}_r$. *Assume that the diagonal elements of* $\mathbf{U}$ *are all positive.*
(4) $\mathbf{A}_n = \mathbf{U}_n\mathbf{L}_n\mathbf{V}_n^* + o(1)$, *where* $\mathbf{U}_n^*\mathbf{U}_n = \mathbf{I}_r + o(1)$, $\mathbf{V}_n^*\mathbf{V}_n = \mathbf{I}_r + o(1)$ *and* $\mathbf{L}_n = diag\{l_{n1}, \ldots, l_{nr}\}$ *is* $r \times r$ *with* $l_{n1} \geqslant \cdots \geqslant l_{nr} \geqslant 0$. *For each column of* $\mathbf{U}_n$, *assume that the first nonzero element on or below the diagonal is positive.*

Then we have $\mathbf{L}_n \to \mathbf{L}$, $\mathbf{U}_n \to \mathbf{U}$ *and* $\mathbf{V}_n \to \mathbf{V}$.

Remark 2. From the proof below, it is clear that the convergence $\mathbf{L}_n \to \mathbf{L}$ does not require distinct singular values of $\mathbf{A}$.

Remark 3. If both $\mathbf{A}_n$ and $\mathbf{A}$ are symmetric (or Hermitian) matrices, similar conclusions hold for eigenvalues. In the proof, one only needs to take $\mathbf{V}_n = \mathbf{U}_n$ and to allow l_{nj} taking negative values.

Proof. The conclusion $\mathbf{L}_n \to \mathbf{L}$ follows from Lemma 1 and condition (2). It then implies that l_{n1} are uniformly bounded. Hence,

$$\mathbf{A}_n\mathbf{A}_n^* = \mathbf{U}_n\mathbf{L}_n^2\mathbf{U}_n^* + o(1).$$

By condition (4), we have

$$\mathbf{A}_n\mathbf{A}_n^*\mathbf{U}_n = \mathbf{U}_n\mathbf{L}_n^2 + o(1). \tag{5}$$

Write $\mathbf{U}_n = [\mathbf{u}_{n1} \vdots \cdots \vdots \mathbf{u}_{nr}]$ and $\mathbf{V}_n = [\mathbf{v}_{n1} \vdots \cdots \vdots \mathbf{v}_{nr}]$. By (5), we have

$$\mathbf{A}_n\mathbf{A}_n^*\mathbf{u}_{nj} = l_{nj}^2\mathbf{u}_{nj}.$$

For $j \leqslant r$, let $\tilde{\mathbf{u}}_j$ be an arbitrary limit point of the bounded sequence $\{\mathbf{u}_{nj}\}$. By condition (2) and the fact that $l_{nj} \to l_j$, we have

$$\mathbf{A}\mathbf{A}^*\tilde{\mathbf{u}}_j = l_j^2\tilde{\mathbf{u}}_j.$$

This shows that $\tilde{\mathbf{u}}_j$ is a unit eigenvector of $\mathbf{A}\mathbf{A}^*$, corresponding to the eigenvalue l_j^2. By condition (1), the dimension of the eigenspace of $\mathbf{A}\mathbf{A}^*$ corresponding to l_j^2 is one. Hence, $\tilde{\mathbf{u}}_j = c\mathbf{u}_j$, where $|c| = 1$ and $\mathbf{u}_j$ is the jth column of $\mathbf{U}$. Since the jth element of $\mathbf{u}_j$ is positive and the jth element of $\tilde{\mathbf{u}}_j$ is nonnegative, we conclude that $c = 1$, or equivalently $\tilde{\mathbf{u}}_j = \mathbf{u}_j$. Consequently, $\mathbf{u}_{nj} \to \mathbf{u}_j$. This proves that $\mathbf{U}_n \to \mathbf{U}$.

By conditions (2)–(4),

$$\mathbf{V}_n^* = \mathbf{L}_n^{-1}(\mathbf{U}_n^*\mathbf{U}_n)^{-1}\mathbf{U}_n^*\mathbf{A}_n + o(1) \to \mathbf{L}^{-1}(\mathbf{U}^*\mathbf{U})^{-1}\mathbf{U}^*\mathbf{A} = \mathbf{V}^*,$$

from which the rest of the lemma follows. □

Lemma 3. *Let $\alpha_n \to 0$ and $\mathbf{A}$ and $\mathbf{A}_n$ be $p \times q$ complex matrices satisfying*

$$\mathbf{A}_n = \Lambda + \alpha_n\mathbf{R} + o(\alpha_n), \tag{6}$$

where $\Lambda = diag\{\lambda_1\mathbf{I}_{p_1}, \dots, \lambda_s\mathbf{I}_{p_s}, 0_{p_{s+1}\times q_{s+1}}\}$, $p_1 + \cdots + p_s = K$, $p_{s+1} = p - K$, $q_{s+1} = q - K$ and $\lambda_1 > \cdots > \lambda_s > 0$. Furthermore, let

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \cdots & \mathbf{R}_{1,s+1} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{s+1,1} & \cdots & \mathbf{R}_{s+1,s+1} \end{pmatrix}$$

where $\mathbf{R}_{gh}$ is a $p_g \times q_h$ matrix $(g, h = 1, \dots, s+1)$ with $q_g = p_g$ for $g = 1, \dots, s$. Let the singular value decomposition of $\mathbf{A}_n$ take the form

$$\mathbf{A}_n = \mathbf{U}_n\mathbf{L}_n\mathbf{V}_n^*$$

where $\mathbf{L}_n = diag[l_{n1}, \dots, l_{n,r}]$, $\mathbf{U}_n^*\mathbf{U}_n = \mathbf{V}_n^*\mathbf{V}_n = \mathbf{I}_r$. Then, we have the following results.

(i) *For $k \leqslant s$,*

$$\{\alpha_n^{-1}(l_{nj} - \lambda_k),\ p_1 + \cdots + p_{k-1} < j \leqslant p_1 + \cdots + p_k\}$$

tends to the set of eigenvalues of $(R_{kk} + R_{kk}^*)/2$, and

$$\{\alpha_n^{-1}l_{nj},\ K < j \leqslant r\}$$

tends to the set of singular values of $R_{s+1,s+1}$.

(ii) *If the* $p \times r$ *matrix* $\mathbf{U}_n = (\mathbf{U}_{nij})_{i,j=1}^{s+1}$ *is split into blocks of the* $p_1, \ldots, p_s, r-K$ *columns and* $p_1, \ldots, p_s, \; p-K$ *rows, and the* $q \times r$ *matrix* $\mathbf{V}_n = (\mathbf{V}_{nij})_{i,j=1}^{s+1}$ *is split into blocks of* $p_1, \ldots, p_s, r-K$ *columns and* $p_1, \ldots, p_s, q-K$ *rows, then for* $i \neq j$,

$$\mathbf{U}_{nij} \to 0, \quad \textit{and} \quad \mathbf{V}_{nij} \to 0.$$

(iii) *If the eigenvalues of* $(\mathbf{R}_{kk} + \mathbf{R}_{kk}^*)/2$, $k \leqslant s$, *are distinct and its eigenmatrix* $\mathbf{U}_{kk}$ *has positive diagonal elements, then*

$$\mathbf{U}_{nkk} \to \mathbf{U}_{kk} \quad \textit{and} \quad \mathbf{V}_{nkk} \to \mathbf{U}_{kk},$$

provided that $\mathbf{U}_{n,kk}$ *is chosen to have nonnegative diagonal elements.*

(iv) *If the singular values of* $\mathbf{R}_{s+1,s+1}$ *are positive and distinct and its left eigenmatrix* $\mathbf{U}_{s+1,s+1}$ *has positive diagonal elements, then*

$$\mathbf{U}_{n,s+1,s+1} \to \mathbf{U}_{s+1,s+1} \quad \textit{and} \quad \mathbf{V}_{n,s+1,s+1} \to \mathbf{V}_{s+1,s+1},$$

provided that $\mathbf{U}_{n,s+1,s+1}$ *is chosen to have nonnegative diagonal elements, where* $\mathbf{V}_{s+1,s+1}$ *is the right eigenmatrix in the singular value decomposition of* $\mathbf{R}_{s+1,s+1}$.

Proof. By Lemma 1, we have $|l_{nj} - \lambda_k| = O(\alpha_n)$ for $p_1 + \cdots + p_{k-1} < j \leqslant p_1 + \cdots + p_k$. We also have

$$\mathbf{U}_n \mathbf{L}_n^2 = \Lambda\Lambda' \mathbf{U}_n + \alpha_n(\mathbf{R}\Lambda' + \Lambda\mathbf{R}^*)\mathbf{U}_n + o(\alpha_n). \tag{7}$$

Note that

$$\mathbf{U}_n = \begin{pmatrix} \mathbf{U}_{n11} & \cdots & \mathbf{U}_{n,1,s+1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{n,s+1,1} & \cdots & \mathbf{U}_{n,s+1,s+1} \end{pmatrix}$$

where the order of $\mathbf{U}_{njk}$ is $p_j \times p_k$ with $p_{s+1} = p - K$.

Comparing both sides of (7), we find that, for $j \neq k$,

$$\mathbf{U}_{n,jk} = O(\alpha_n)$$

which, together with the fact that $\mathbf{U}_{n1,k}^* \mathbf{U}_{n1,k} + \cdots + \mathbf{U}_{n,s+1,k}^* \mathbf{U}_{n,s+1,k} = \mathbf{I}_{p_k}$, implies that, for $j = k = 1, \ldots, s$,

$$\mathbf{U}_{nkk}^* \mathbf{U}_{nkk} = \mathbf{I}_{p_k} + O(\alpha_n^2) \quad \text{and} \quad \mathbf{U}_{n,s+1,s+1}^* \mathbf{U}_{n,s+1,s+1} = \mathbf{I}_{r-K} + O(\alpha_n^2).$$

Also, by (7), we have

$$\mathbf{U}_{nkk}[\alpha_n^{-1}(\mathbf{L}_{nk}^2 - \lambda_k^2 \mathbf{I}_{p_k})] = \lambda_k(\mathbf{R}_{kk} + \mathbf{R}_{kk}^*)\mathbf{U}_{n,kk} + o(1)$$

for $k \leqslant s$. Note that the left-hand side of above is asymptotically equivalent to $2\lambda_k \mathbf{U}_{nkk}[\alpha_n^{-1}(\mathbf{L}_{nk} - \lambda_k \mathbf{I}_{p_k})]$. Then by Remark 3, we know that $\alpha_n^{-1}(\mathbf{L}_{nk} - \lambda_k \mathbf{I}_{p_k})$ tends to the diagonal matrix consisting of the eigenvalues of $(\mathbf{R}_{kk} + \mathbf{R}_{kk}^*)/2$. Furthermore, if the eigenvalues of the matrix $(\mathbf{R}_{kk} + \mathbf{R}_{kk}^*)/2$ are distinct and the diagonal elements of $\mathbf{U}_{kk}$ and $\mathbf{U}_{nkk}$ are chosen to have positive diagonal elements, then $\mathbf{U}_{nkk} \to \mathbf{U}_{kk}$.

Similarly, we have

$$\mathbf{V}_n \mathbf{L}_n^2 = \Lambda'\Lambda \mathbf{V}_n + \alpha_n(\mathbf{R}^*\Lambda + \Lambda'\mathbf{R})\mathbf{V}_n + o(\alpha_n). \tag{8}$$

From (8), we obtain for $i \neq j$,

$$\mathbf{V}_{n,ij} = O(\alpha_n).$$

By multiplying $\mathbf{V}_n$ to both sides of the equation $\mathbf{U}_n\mathbf{L}_n\mathbf{V}_n^* = \Lambda + O(\alpha_n)$ and extracting the kth diagonal block for $k \leqslant s$, we obtain

$$\mathbf{U}_{nkk}\mathbf{L}_{nk} = \lambda_k\mathbf{V}_{nkk} + O(\alpha_n).$$

It then follows that $\mathbf{V}_{nkk} \to \mathbf{U}_{kk}$.

Finally, consider the case $k = s + 1$. We have from (6)

$$\alpha_n^{-1}\mathbf{U}_{n,s+1,s+1}\mathbf{L}_{n,s+1}\mathbf{V}^*_{n,s+1,s+1} = \mathbf{R}_{s+1,s+1} + o(1).$$

By Lemma 2, $\alpha_n^{-1}l_{nj}$ $(K < j \leqslant r)$ tend to the set of singular values of $R_{s+1,s+1}$ and the conclusion (iv) also follows. All parts of Lemma 3 are now proven. $\square$

References

[1] T.W. Anderson, Multivariate Analysis, Wiley, New York, 1984.
[2] Z.D. Bai, Methodologies in spectral analysis of large-dimensional random matrices, a review (with discussions), Statist. Sin. 9 (1999) 611–677.
[3] R.D. Cook, Principal Hessian directions revisited, J. Amer. Statist. Assoc. 93 (1998) 84–100.
[4] R.D. Cook, X. Yin, Dimension reduction and visualization in discriminant analysis (with discussion), Aust. NZ J. Statist. 43 (2001) 147–199.
[5] N. Duan, K.C. Li, Slicing regression: a link-free regression method, Ann. Statist. 19 (1991) 505–530.
[6] M.L. Eaton, D. Tyler, The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis, J. Multivariate Anal. 50 (1994) 238–264.
[7] K.Y. Fan, Maximum properties and inequalities for the eigenvalues of completely continuous operators, Proc. Natl Acad. Sci. USA 37 (1951) 760–766.
[8] W.K. Fung, X. He, L. Liu, P.D. Shi, Dimension reduction based on canonical correlation, Statist. Sin. 12 (2002) 1093–1114.
[9] X. He, L.J. Shen, Linear regression after spline transformation, Biometrika 84 (1997) 474–481.
[10] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–327.
[11] S. Portnoy, On the central limit theorem in R^p when $p \to \infty$, Probab. Theory Related Fields 73 (1986) 571–583.
[12] S.R. Searle, Linear Models, 1st Edition, Wiley, New York, 1971.